

生成AI(人工知能)発の金融リスク

生成AIの急速な進展のなか、金融機関においてもその活用が進んでいる一方、そこには、金融セクターならではのリスクも存在する。今後は、公平性や安全性といったAI固有のリスクを極小化しつつ、AIの活用を通じて利便性や革新性を追求する「責任あるAI」の開発・提供が重要となる。

谷口 栄治

調査部金融リサーチセンター
主任研究員

金融分野における 生成AIの活用と期待される効果

生成AI(人工知能)は、自ら新たなコンテンツを生成できる点が特徴であり、2022年11月にOpenAI社が対話型AIのChatGPTを公表して以降、爆発的に普及している。金融分野においても、生成AI等の活用による効果として、①業務効率化に伴うコスト削減、②新たな商品・サービスの開発サポート、③顧客利便性・満足度の向上、④与信判断能力の高度化、⑤複雑なデータ解析による投資判断の改善、⑥リスク管理やコンプライアンス態勢の高度化、などが期待されている。

想定される「生成AI発」の 金融リスク

一方、金融分野におけるAIの活用によって、新たなリスクの温床になるとの指摘も多く、大別すれば、以下の項目に集約される(図表)。

データプライバシー・セキュリティ

金融機関は、顧客の個人情報や企業秘密、資産状況や入金データなど、極めて機密性の高い情報を取り扱っている。生成AIを利用するにあたり、そのようなデータや情報を入力したり、学習に用いたりする際、機微な情報が漏洩したり、推測されたりするリスクがある。そのため、データセキュリティやプライバシー保護の観点から、国内外の多くの金融機関で社内シス

テムからのChatGPT等への直接的なアクセスが禁じられており、IT企業と連携して自社専用仕様として開発したAIシステムが導入されている。

埋め込まれた偏見・バイアス

学習するデータに偏りがあれば、意思決定・判断にバイアスが生じ、金融排除や社会的信頼の毀損につながるおそれがある。例えば、個人向けローンの与信判断にAIを活用するケースで、性別や人種、居住地等によって借入条件に差異が生じたり、マイノリティ層が必要な融資を受けられなかったりするなど、差別の助長や不当な金融排除のリスクが指摘されている。

金融インフラとしての頑健性・確実性

生成AIでは、質問に対して流暢な言語で返信される点が評価される一方、事実と異なる情報を生成することがある。このように生成AIが、誤った情報を、さも本当であるかのように回答する事態は、幻覚(ハルシネーション)と呼ばれる。金融分野では、顧客への情報提供や金融機関のリスク評価等において誤情報が提供されることがあれば、顧客の意思決定が歪められたり、金融機関がリスクテイクやリスク管理において誤った判断を下したりして、金融システムや消費者保護に悪影響が生じるおそれがある。

説明可能性・説明責任(アカウントビリティ)

金融機関は、当局を含めさまざまなステークホルダーに対して、自らの意思決定や行動に関する説明責任を有する。もっとも、AIアルゴリズムの結果として下される意思決定や行動を説明することは困難を伴う

図表 AI活用にあたっての金融関連の主なリスク

	概要
データプライバシー・セキュリティ	入力や学習に用いられる機微なデータ・情報が漏えいしたり、推測されたりしないか
埋め込まれた偏見・バイアス	意思決定・判断にバイアスが生じ、金融排除や社会的信頼の毀損につながらないか
金融インフラとしての頑健性・確実性	誤情報が提供され、顧客や金融機関の意思決定や判断を誤らせることはないか
説明可能性・説明責任	AI内の意思決定プロセスがブラックボックス化し、説明責任を果たせなくなるか
サイバーセキュリティ・悪用リスク	金融犯罪や相場操縦等に悪用されないか（フィッシング詐欺、ディープフェイク等）
金融システムの安全性	リスク評価やリスク判断が自動化・画一化され、金融市場の変動（プロシクリカリティ）が増幅・加速し、システミックリスクにつながらないか

出所：IMF「Generative Artificial Intelligence in Finance: Risk Considerations」等を基に日本総研作成

ほか、生成AIでは、アルゴリズムが複雑化しており、意思決定がブラックボックス化するおそれがある。説明責任を果たせない状況が続けば、金融機関としての信頼性やブランドイメージを毀損したり、金融商品やサービスに対する信頼性が損なわれたりするリスクがある。

サイバーセキュリティ・悪用リスク

生成AIが悪用され、金融犯罪や相場操縦等の温床になるリスクが存在する。例えば、フィッシング詐欺等に用いられるメール等の文面が極めて自然なものとなり、被害が拡大するおそれがある。また、ディープフェイク（フェイク画像、フェイク動画）が拡散することで、金融市場が変動したり、金融機関や投資家の投資行動を誤らせたりするリスクも指摘されている。実際、2023年5月には、米国防総省（ペンタゴン）の近くで爆発が起きたとする偽画像がSNS上で拡散され、株価が一時下落するといった事態があった。さらに、生成AIの学習データに悪意ある情報やデータが混入されるなどのサイバー攻撃を受け、アルゴリズムが改ざんされるといった問題もある。

金融システムの安定性

生成AIが広く普及し、それを利用する投資家や金融機関が増えれば増えるほど、投資判断やリスク評価のプロセスが自動化、画一化され、金融市場の変動（プロシクリカリティ）が増幅されるリスクが指摘されている。具体的には、投資家が生成AIを用いて投資を行う際に、共通のアルゴリズムが用いられることで、多

くの投資家が同じ投資行動を取り、結果的に金融市場が一方向に大きく振れるおそれがある。また、生成AIのプロバイダー（IT企業）が独占的な立場にあれば、投資行動がさらに画一化し、金融システムの安定性に問題が生じるリスクが高まる。

わが国金融セクターに求められる対応

生成AIの普及が進むなかで今後重要となるのが、「責任あるAI」、つまり公平性や安全性といったAI固有のリスクを極小化しつつ、AIの活用を通じて利便性や革新性を追求するという姿勢である。わが国金融セクターとしては、AIの活用による新たな付加価値の提供に努めていくとともに、金融システムの健全性、顧客保護の観点からも、内外の金融機関や当局と連携しながらAIのリスクを適切に管理・監督していくことが求められる。X

Profile

谷口 栄治

(たにぐち・えいじ)

2007年三井住友銀行入行。2010年7月から2012年6月、経済産業省経済産業政策局調査課。2012年7月から、三井住友銀行経営企画部金融調査室。2020年4月より、日本総研調査部金融リサーチセンター。

